(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/045990 A2

(51) International Patent Classification[7]: C07K 14/46, 14/47, C12N 1/16, 1/21, 5/10, 15/12, 15/63, C12Q 1/68, G01N 33/53

(21) International Application Number: PCT/EP02/13866

(22) International Filing Date:
26 November 2002 (26.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/333,348	26 November 2001 (26.11.2001)	US
60/384,537	31 May 2002 (31.05.2002)	US
60/422,471	30 October 2002 (30.10.2002)	US

(71) Applicant *(for all designated States except US)*: **HYBRIGENICS** [FR/FR]; 3/5 Impasse Reille, F-75014 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEGRAIN, Pierre** [FR/FR]; 5, rue Mizon, F-75015 Paris (FR). **GAUTHIER, Jean-Michel** [FR/FR]; 9 Avenue du Maréchal Galliéni, F-78700 Conflans Sainte Honorine (FR). **COLLAND, Frédéric** [FR/FR]; 16, rue du Manoir, F-95380 Puiseux (FR). **JACQ, Xavier** [FR/FR]; 11 rue de Rambouillet, F-75012 Paris (FR).

(74) Agents: **ERNEST Gutmann-Yves Plasseraud S.A.** et al.; 3 rue Chauveau-Lagarde, F-75008 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/045990 A2

(54) Title: PROTEIN-PROTEIN INTERACTIONS INVOLVING TRANSFORMING GROWTH FACTOR β SIGNALING OR INVOLVING TRANSDUCTION SIGNALS OF TRANSFORMING FACTOR β FAMILY MEMBERS

(57) Abstract: The present invention relates to protein-protein interactions involved in transforming growth factor β disorders and/or diseases. More specifically, the present invention relates to complexes of polypeptides or polynucleotides encoding the polypeptides, fragments of the polypeptides, antibodies to the complexes, Selected Interacting Domains (SID®) which are identified due to the protein-protein interactions, methods for screening drugs for agents which modulate the interaction of proteins and pharmaceutical compositions that are capable of modulating the protein-protein interactions.

PROTEIN-PROTEIN INTERACTIONS INVOLVING TRANSFORMING GROWTH FACTOR β SIGNALING OR INVOLVING TRANSDUCTION SIGNALS OF TRANSFORMING FACTOR β FAMILY MEMBERS

The present application claims priority to US provisional applications No. 60/333,348 filed on November 26, 2001, No. 60/384,537 filed on May 31, 2002 and No. 60/422,471 filed on October 30, 2002.

BACKGROUND AND PRIOR ART

Most biological processes involve specific protein-protein interactions. Protein-protein interactions enable two or more proteins to associate. A large number of non-covalent bonds form between the proteins when two protein surfaces are precisely matched. These bonds account for the specificity of recognition. Thus, protein-protein interactions are involved, for example, in the assembly of enzyme subunits, in antibody-antigen recognition, in the formation of biochemical complexes, in the correct folding of proteins, in the metabolism of proteins, in the transport of proteins, in the localization of proteins, in protein turnover, in first translation modifications, in the core structures of viruses and in signal transduction.

General methodologies to identify interacting proteins or to study these interactions have been developed. Among these methods are the two-hybrid system originally developed by Fields and co-workers and described, for example, in U.S. Patent Nos. 5,283,173, 5,468,614 and 5,667,973, which are hereby incorporated by reference.

The earliest and simplest two-hybrid system, which acted as basis for development of other versions, is an *in vivo* assay between two specifically constructed proteins. The first protein, known in the art as the "bait protein" is a chimeric protein which binds to a site on DNA upstream of a reporter gene by means of a DNA-binding domain or BD. Commonly, the binding domain is the DNA-binding domain from either Gal4 or native *E. coli* LexA and the sites placed upstream of the reporter are Gal4 binding sites or LexA operators, respectively.

The second protein is also a chimeric protein known as the "prey" in the art. This second chimeric protein carries an activation domain or AD. This activation domain is typically derived from Gal4, from VP16 or from B42.

Besides the two-hybrid systems, other improved systems have been developed to detected protein-protein interactions. For example, a two-hybrid plus one system was developed that allows the use of two proteins as bait to screen available cDNA libraries to detect a third partner. This method permits the detection between proteins that are part of a larger protein complex such as the RNA polymerase II holoenzyme and the TFIIH or TFIID complexes. Therefore, this method, in general, permits the detection of ternary complex

formation as well as inhibitors preventing the interaction between the two previously defined fused proteins.

Another advantage of the two-hybrid plus one system is that it allows or prevents the formation of the transcriptional activator since the third partner can be expressed from a conditional promoter such as the methionine-repressed Met25 promoter which is positively regulated in medium lacking methionine. The presence of the methionine-regulated promoter provides an excellent control to evaluate the activation or inhibition properties of the third partner due to its "on" and "off" switch for the formation of the transcriptional activator. The three-hybrid method is described, for example in Tirode *et al.*, *The Journal of Biological Chemistry*, **272**, No. 37 pp. 22995-22999 (1997) incorporated herein by reference.

Besides the two and two-hybrid plus one systems, yet another variant is that described in Vidal et al, *Proc. Natl. Sci. 93* pgs. 10315-10320 called the reverse two- and one-hybrid systems where a collection of molecules can be screened that inhibit a specific protein-protein or protein-DNA interactions, respectively.

A summary of the available methodologies for detecting protein-protein interactions is described in Vidal and Legrain, *Nucleic Acids Research* Vol. 27, No. 4 pgs. 919-929 (1999) and Legrain and Selig, FEBS Letters 480 pgs. 32-36 (2000) which references are incorporated herein by reference.

However, the above conventionally used approaches and especially the commonly used two-hybrid methods have their drawbacks. For example, it is known in the art that, more often than not, false positives and false negatives exist in the screening method. In fact, a doctrine has been developed in this field for interpreting the results and in common practice an additional technique such as co-immunoprecipitation or gradient sedimentation of the putative interactors from the appropriate cell or tissue type are generally performed. The methods used for interpreting the results are described by Brent and Finley, Jr. in *Ann. Rev. Genet., 31 pgs. 663-704 (1997).* Thus, the data interpretation is very questionable using the conventional systems.

One method to overcome the difficulties encountered with the methods in the prior art is described in WO99/42612, incorporated herein by reference. This method is similar to the two-hybrid system described in the prior art in that it also uses bait and prey polypeptides. However, the difference with this method is that a step of mating at least one first haploid recombinant yeast cell containing the prey polypeptide to be assayed with a second haploid recombinant yeast cell containing the bait polynucleotide is performed. Of course the person skilled in the art would appreciate that either the first recombinant yeast cell or the second recombinant yeast cell also contains at least one detectable reporter gene that is activated by a polypeptide including a transcriptional activation domain.

The method described in WO99/42612 permits the screening of more prey polynucleotides with a given bait polynucleotide in a single step than in the prior art systems due to the cell to cell mating strategy between haploid yeast cells. Furthermore, this method is more thorough and reproducible, as well as sensitive. Thus, the presence of false negatives and/or false positives is extremely minimal as compared to the conventional prior art methods.

Transforming growth factor β (TGFβ) belongs to a super-family of cytokines, including TGFβ1, TGFβ2, TGFβ3, activins and Bone Morphologenetic Proteins (hereinafter BMP), which are synthesized by many cell types and have a variety of cellular and biological effects, including control of proliferation, differentiation, migration, angiogenesis, immunity and regulation of the turnover of the extracellular matrix. A number of disease states are known to be associated with variations in expression of genes which are controlled by TGFβ and related ,cytokines, including fibrotic disorders, abnormal wound healing, abnormal bone formation, cancer and tumor development, neurologic disorders, haematopoiesis and immune and inflammatory disorders.

Signaling by this family of cytokines is transduced by heteromeric complexes of transmembrane Ser/Thr kinase receptors. Upon ligand binding, type II receptor phosphorylates and activates type I receptor which then propagates signals to downstream targets, in particular the Smad proteins.

Ten mammalian Smad proteins have been identified and divided into three classes. The first includes pathway-restricted proteins such as Smad1, Smad5 and Smad8 which are specifically involved in BMP signaling and Smad2 and Smad3 which are restricted to TGFβ/activin pathway. The second class contains the common-mediator Smad4 implicated in both BMP and TGFβ/activin pathways. The third class contains the inhibitory Smads, Smad6 and Smad7. At least Smad2 and Smad3 are retained in the cytoplasm by binding to the SARA protein. After phosphorylation by TGFβ-activated type I receptor on their carboxy-terminal SSXS sequence, pathway-restricted Smads form heteromeric complexes with Smad4 and then translocate to the nucleus where they control expression of diverse genes involved in various biological processes such as control of cellular proliferation and differentiation, regulation of the immune system and regulation of the extracellular matrix formation.

Several proteins such as TGIF, Ski, SnoN, SNIP1 and CBP have been identified as Smad transcriptional co-regulators and shown to modulate the transcriptional ability of Smad proteins by direct interactions. Finally, proteins such Smurf1 and Smurf2 are involved in degradation of Smad proteins by the proteasome machinery.

Most biological processes involve specific protein-protein interactions. Protein-protein interactions enable two or more proteins to associate. A large number of non-covalent bonds

form between the proteins when two protein surfaces are precisely matched. These bonds account for the specificity of recognition. Thus, protein-protein interactions are involved, for example, in the assembly of enzyme subunits, in antibody-antigen recognition, in the formation of biochemical complexes, in the correct folding of proteins, in the metabolism of proteins, in the transport of proteins, in the localization of proteins, in protein turnover, in first translation modifications, in the core structures of viruses and in signal transduction.

Several members of the TGFβ/BMP pathways (SARA, Smurf1, Smurf2, Smad1, Smad2/hMAD2, Smad3/hMAD-3, Smad4, Smad5/MADH5, Smad7, Smad9/MADH6, SNIP1, SnoN) have been used as baits in yeast-two hybrid screening experiments. Several proteins have been identified as interactors with thoses baits (Figure 10). It was showed here functional data in mammalian cells that validate that those interactants are proteins involved in TGFβ/BMP signaling.

Thus, there is the still a need to explore all mechanisms relating to transforming growth factor β protein and to identify drug targets for fibrotic disorders, abnormal wound healing, abnormal bone formation, cancer and tumor development, neurologic disorders, haematopoiesis and immune and inflammatory disorders and/or diseases.

SUMMARY OF THE PRESENT INVENTION

Thus, it is an aspect of the present invention to identify protein-protein interactions involving proteins of the transforming growth factor β super-family of cytokines transduction pathway and to identify drug targets for fibrotic disorders, abnormal wound healing, abnormal bone formation, cancer and tumor development, neurologic disorders, haematopoiesis and immune and inflammatory disorders and/or disease.

It is another aspect of the present invention to identify protein-protein interactions involved in transforming growth factor β-mediated disorders and/or diseases for the development of more effective and better targeted therapeutic treatments.

It is yet another aspect of the present invention to identify complexes of polypeptides or polynucleotides encoding the polypeptides and fragments of the polypeptides of the transforming growth factor β super-family of cytokines transduction pathway.

It is yet another aspect of the present invention to identify antibodies to these complexes of polypeptides or polynucleotides encoding the polypeptides and fragments of the polypeptides involving transforming growth factor β signaling including polyclonal, as well as monoclonal antibodies that are used for detection.

It is still another aspect of the present invention to identify selected interacting domains of the polypeptides, called SID® polypeptides.

It is still another aspect of the present invention to identify selected interacting domains of the polynucleotides, called SID® polynucleotides.

It is still another aspect of the present invention to provide a diagnostic kit to test for deficiencies in the transforming growth factor β super-family of cytokines transduction pathway.

It is another aspect of the present invention to identify interacting proteins in the transforming growth factor β super-family of cytokines transduction pathway that can be used in pharmaceutical compositions or for diagnostic purposes.

It is another aspect of the present invention to generate protein-protein interactions maps called PIM®s.

It is yet another aspect of the present invention to provide a method for screening drugs for agents which modulate the interaction of proteins and pharmaceutical compositions that are capable of modulating the protein-protein interactions involved in transforming growth factor β disorders and/or diseases.

It is another aspect to administer the nucleic acids of the present invention via gene therapy.

It is yet another aspect of the present invention to provide protein chips or protein microarrays.

It is yet another aspect of the present invention to provide a report in, for example paper, electronic and/or digital forms, concerning the protein-protein interactions, the modulating compounds and the like as well as a PIM®.

These and other aspects are achieved by the present invention as evidenced by the summary of the invention, description of the preferred embodiments and the claims.

Thus the present invention relates to a complex of interacting proteins of columns 1 and 4 of Table 2.

Furthermore, the present invention provides SID® polynucleotides and SID® polypeptides of Table 3, as well as a PIM® involved in transforming growth factor β-mediated disorders and/or diseases.

The present invention also provides antibodies to the protein-protein complexes involved in transforming growth factor β-mediated disorders and/or diseases.

In another embodiment the present invention provides a method for screening drugs for agents that modulate the protein-protein interactions and pharmaceutical compositions that are capable of modulating protein-protein interactions.

In another embodiment the present invention provides protein chips or protein microarrays.

In yet another embodiment the present invention provides a report in, for example, paper, electronic and/or digital forms.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic representation of the pB6 plasmid.

Fig. 2 is a schematic representation of the pB20 plasmid.

Fig. 3 is a schematic representation of the pP6 plasmid.

Fig. 4 is a schematic representation of vectors expressing the T25 fragment.

Fig. 5 is a schematic representation of vectors expressing the T18 fragment.

Fig. 6 is a schematic representation of various vectors of pCmAHL1, pT25 and pT18.

Fig. 7 is a schematic representation identifying the SID®'s of proteins of the present invention. In this figure the "Full-length prey protein" is the Open Reading Frame (ORF) or coding sequence (CDS) where the identified prey polypeptides are included. The Selected Interaction Domain (SID®) is determined by the commonly shared polypeptide domain of every selected prey fragment.

Fig. 8 is a protein map (PIM®).

Fig. 9 is a schematic representation of the pB27 plasmid.

Fig. 10 is a schematic representation of the pB28 plasmid.

Fig. 11 is a schematic representation of a protein interaction map around the newly functionally characterized proteins described in the present invention. These 10 proteins are highlighted by the symbol "*". The Predicted Biological Score (PBS) is represented by a code on each line and classified from A to E (Rain *et al.*, 2001). PP1ca is also named PPP1CA. MADH5 and MADH6 correspond to Smad5 and Smad9, respectively. hMAD-2 and h-MAD-3 correspond to Smad2 and Smad3, respectively. MAN1 is the orthologous of SANE, a protein recently identified as involved in the BMP pathway (Raju *et al.*, 2002)

Fig. 12 is a schematic representation of a protein interaction map between ZNF8 and Smad proteins. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, ZNF8 was shown to interact with Smad1 (A), Smad4 (B), Smad5 (C) and Smad9 (D). Amino-acid position are indicated.

Fig. 13 A, B and C are graphs showing that ZNF8 siRNA represses TGFβ- and BMP-dependent luciferase reporter activities. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (A & B) or the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (C). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGFβ pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of ZNF8-targeting siRNA duplex was tested in cells treated or not with 50ng/ml BMP7 (A), 50ng/ml BMP6 (B) or 5 ng/ml TGFβ1 (C) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or

three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 14A, B and C are graphs showing that ZNF8 siRNA specifically represses BMP-dependent markers. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or ZNF8-targeting siRNA duplex. Cells were treated or not with 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed either at 0.5nM and 2.5nM (A & B) or at 4 and 40nM (C) of duplex. Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the BMP pathway markers junB (A) and alkaline phosphatase (B& C). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using either GAPDH (A & B) or hGUS (C).

Fig. 15 A and B are graphs showing that ZNF8 siRNA does not repress BMP-independent markers. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or ZNF8-targeting siRNA duplex. Cells were treated or not with 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed either at 0.5nM and 2.5nM (A) or at 4 and 40nM (B) of duplex. Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the TGFβ pathway marker PAI-1 (PAI-1 hereinafter Plasminogen Activator Inhibitor I) (A) and an unrelated marker, hGUS (B). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using either GAPDH (A) or relative levels (B).

Fig. 16 is a schematic representation of an Interaction between LAPTm5 and Smurf2. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, interaction between Smurf2 and LAPTm5 was found in both directions. Smurf2 was shown to interact with the C-terminal domain of LAPTm5.

Fig. 17 A and B are graphs showing that LAPTm5 specifically inhibits the TGFβ pathway. The effect of LAPTm5 over-expression was studied using the following Luciferase reporter vectors: a TGFβ responsive element (TGF-RE = $p(GTCT)_6$-MLP-Luc), a BMP-responsive element (BMP-RE = $p(GC)_{12}$-MLP-Luc) and an unrelated reporter (pGL3 control) (see Materials & Methods). The effect was studied in the presence or absence of TGFβ (10 ng/ml) or BMP7 (50 ng/ml), as described. This study was performed with 0, 2 or 10 ng of

pV3-LAPTm5 in HepG2 cells (A) or with 0, 0.5, 2, 10 or 50 ng of pV3-LAPTm5 in HEK293 cells (B). The specific Luciferase activity was normalized using the pRL-TK vector. Experiments were performed in triplicate.

Fig. 18 A and B are graphs showing that LAPTm5 expression is up-regulated by TGFβ The endogenous level of LAPTm5 mRNA was determined in several cell lines by Q-PCR experiments using the LAPTm5 probe (see Materials & Methods). Ct levels of LAPTm5 mRNA is given for each cell lines (A). The endogenous level of mRNA was determined in HepG2 cells in the presence or absence of TGFβ (10 ng/ml) with or without a TβRI-targeting siRNA duplex (B) (TβRI hereinafter Transforming Growth Factor β Receptor I.

Fig. 19 A and B are graphs showing that LAPTm5 siRNA up-regulates BMP and TGFβ-dependent reporter activities. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the TGFβ reponsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (A) or the BMP responsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (B). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGFβ pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of LAPTm5-targeting siRNA duplex was tested in cells treated or not with 5ng/ml recombinant human TGFβ (A), 50ng/ml recombinant human BMP7 (B) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 20 A, B, C and D are graphs showing that LAPTm5 siRNA up-regulates BMP and TGFβ-dependent markers. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or LAPTm5-targeting siRNA duplex. Cells were treated or not with 5 ng/ml of recombinant human TGFβ1 or 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 40nM of duplex (A, B, C & D). Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the TGFβ pathway markers PAI-1 and junB (A & B, respectively) and a BMP pathway marker, alkaline phosphatase (C). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using hGUS (A, B & C). Relative levels of hGUS in the same experiment are also shown (D).

Fig. 21 is a schematic representation of an Interaction between RNF11 Smurf1, Smurf2 and SARA.The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, RNF11 was shown to interact with Smurf1 (A), Smurf2 (B), and SARA (C). Amino-acid positions are indicated.

Fig. 22 is a gel showing that RNF11 is involved in regulating SARA protein levels. Transfection experiments with pV3-SARA (200 ng) and/or pV3-RNF11 (300 ng) in the presence or absence of TGFβ (10 ng/ml) were performed. After TGFβ induction for 18H, cells' lysates were resolved on a 4-12% NuPAGE gradient gel, transfered and revealed using anti-SARA antibody (see Materials & Methods).

Fig. 23 is a schematic diagram showing the Interaction between KIAA1196 and Smad1. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, KIAA1196 was shown to interact with Smad1.

Fig. 24 A and B are graphs showing that KIAA1196 siRNA specifically represses TGFβ-dependent markers in HepG2 cells.HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (A) or the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (B). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGFβ pathway. A mutated version of the T RI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of KIAA1196-targeting siRNA duplex was tested in cells treated or not with 5ng/ml recombinant human TGFβ (A) and 50ng/ml recombinant human BMP6 (B) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 25 is a graph showing that KIAA1196 siRNA specifically represses TGFβ-dependent reporter activity in HEK293 cells. HEK 293 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc. All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGFβ pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 30nM. Co-transfection of KIAA1196-targeting siRNA duplex was tested in cells treated or not with 5ng/ml recombinant human TGFβ for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of

lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 26 A, B, C and D are graphs showing that KIAA1196 siRNA specifically represses TGFβ-dependent markers. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or KIAA1196-targeting siRNA duplex. Cells were treated or not with 5 ng/ml of recombinant human TGFβ1 or 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 40nM of duplex (A, B, C & D). Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the TGFβ pathway markers PAI-1 and junB (A & B, respectively) and a BMP pathway marker, alkaline phosphatase (C). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using hGUS (A, B & C). Relative levels of hGUS in the same experiment are also shown (D).

Fig. 27 is a schematic representation showing the Interaction between LMO4 and Smad9. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, LMO4 was shown to interact with Smad9.

Fig. 28 A, B and C are graphs showing that LMO4 siRNA specifically repress a BMP-dependent luciferase reporter. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (A) or the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (B). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGF pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of LMO4-targeting siRNA duplex was tested in cells treated or not with 50ng/ml recombinant human BMP7 or BMP6 (A & B, respectively) and 5ng/ml recombinant human TGFβ (C) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are. representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 29 A and B are graphs showing that LMO4 siRNA specifically represses BMP-induced markers in BMP7-treated HepG2 cells.HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or LMO4-targeting siRNA duplex. Cells were treated or not with 50ng/ml of

recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 0.5 or 2.5nM of duplex (A) and 4 or 40nM of duplex (B). Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the BMP pathway marker alkaline phosphatase (A & B). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using hGUS (A, B).

Fig. 30 A, B and C are graphs showing that LMO4 siRNA does not repress BMP-independent markers in BMP7-treated HepG2 cells.HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or LMO4-targeting siRNA duplex. Cells were treated or not with 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 4 or 40nM of duplex (A, B) and 0.5 or 2.5nMof duplex (C). Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the TGFβ and BMP pathways marker junB (A) and a TGFβ pathway marker, PAI-1 (C). Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using hGUS (A) or using GAPDH (C). Relative levels of hGUS in the same experiment are also shown (B).

Fig. 31 is a schematic diagram showing the interaction between PP1ca and SARA. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, PP1ca was shown to interact with SARA.

Fig. 32 A and B are graphs showing that PP1ca stimulates the TGFβ pathway. The effect of PP1ca over-expression was studied using the following luciferase reporter vectors: a TGFβ responsive element (TGF-RE = $p(GTCT)_8$-MLP-Luc), a BMP-responsive element (BMP-RE = $p(GC)_{12}$-MLP-Luc) and an unrelated reporter (pGL3 control) (see Materials & Methods). The effect was studied in the presence or absence of TGFβ (10 ng/ml) or BMP7 (50 ng/ml), as described. This study was performed with 0, 10, 50 or 200 ng of pV3-PP1ca in HepG2 cells (A) or in HEK293 cells (B). The specific Luciferase activity was normalized using the pRL-TK vector. Experiments were performed in triplicate.

Fig. 33 A, B and C are graphs showing that PP1ca stimulates PAI-1 mRNA expression. Baculoviruses containing the Smad3 or PP1ca genes under the control of the CMV promoter were generated and used to infect HepG2 cells (see Materials & Methods). The over-expression level was checked and quantified by Q-PCR (A). The endogenous PAI-1 mRNA levels were measured by Q-PCR 24 hours post infection with Smad3 or PP1ca-containing

baculoviruses in the presence or absence of TGFβ (10 ng/ml). The value 1 is attributed to the mRNA amount of PAI-1 in the absence of TGFβ and in the absence of infection (B).

Fig. 34 is a schematic diagram showing the Interaction between HYPA and Smad4. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, HYPA was shown to interact with Smad4.

Fig. 35 A, B and C are graphs showing that HYPA siRNA specifically represses BMP-dependent reporter activity.HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (A & B) or the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (C). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGFβ pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of HYPA-targeting siRNA duplex was tested in cells treated or not with 50ng/ml recombinant human BMP7 or BMP6 (A & B, respectively) and 5ng/ml recombinant human TGFβ (C) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 36 is a graph showing that HYPA siRNA represses BMP-dependent markers. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or HYPA-targeting siRNA duplex. Cells were treated or not with 50ng/ml of recombinant human BMP7 for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 0.5 or 2.5nM of duplex. Cells were harvested and lysed 48 hours after transfection. Total RNA were extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the BMP pathway marker alkaline phosphatase. Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using GAPDH.

Fig. 37 is a schematic diagram showing the Interaction between FLJ20037 and SARA. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, FLJ20037 was shown to interact with SARA.

Fig. 38 A, B and C are graphs showing that FLJ20037 stimulates PAI-1 mRNA expression. Baculoviruses containing the Smad3 or FLJ20037 genes under the control of the CMV promoter were generated and used to infect HepG2 cells (see Materials & Methods). The over-expression level was checked and quantified by Q-PCR (A). The endogenous PAI-

1 mRNA levels were measured by Q-PCR 24 hours post infection with Smad3 or FLJ20037-containing baculoviruses in the presence or absence of TGFβ (10 ng/mL). The value 1 is attributed to the mRNA amount of PAI-1 in the absence of TGFβ and in the absence of infection (B).

Fig. 39 is a graph showing that FLJ20037 siRNA down-regulates TGFβ-dependent markers.HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with a control siRNA (pGL3-targeting siRNA) or FLJ20037-targeting siRNA duplex. Cells were treated or not with 5ng/ml of recombinant human TGFβ for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. SiRNA transfections were performed at 0.5 or 2.5nM of duplex. Cells were harvested and lysed 48 hours after transfection. Total RNA was extracted as described under *Materials &* Methods and quantitative PCR analysis were performed in order to quantitate the endogenous levels of the TGFβ pathway marker PAI-1. Data are representative of two or three independant duplicated experiments and are presented as normalized RNA levels using GAPDH.

Fig. 40 is a schematic diagram showing the Interaction between PTPN12 and Smad5. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, PTPN12 was shown to interact with Smad5. Amino-acid positions are indicated.

Fig. 41 A and B are graphs showing that PTPN12 siRNA up-regulates BMP and TGFβ-dependent reporter activities. HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (A) or the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (B). All experiments included pRL-TK as an internal transfection control. A TβRI-targeting siRNA duplex was used as a positive control for disruption of the TGF pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of PTPN12-targeting siRNA duplex was tested in cells treated or not with 50ng/ml recombinant human BMP6 (A) and 5ng/ml recombinant human TGFβ (B) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

Fig. 42 A and B are schematic diagrams showing the Interaction between HIPK3, SnoN and SNIP1. The full-length proteins are represented in grey and black boxes correspond to the interaction domains. Using two-hybrid screening, HIPK3 was shown to interact with the N-terminal domains of SNIP1 (A) and SnoN (B). Amino-acid positions are indicated.

Fig. 43 A and B are graphs showing that HIPK3 siRNA specifically up-regulates BMP-dependent reporter activities.

HepG2 cells were transiently transfected in 24 well-plates as described under *Materials & Methods* with the BMP reponsive luciferase reporter, $p(GC)_{12}$-MLP-Luc (A) or the TGFβ responsive luciferase reporter, $p(GTCT)_8$-MLP-Luc (B). All experiments included pRL-TK as an internal transfection control. A T RI-targeting siRNA duplex was used as a positive control for disruption of the TGF pathway. A mutated version of the TβRI-targeting siRNA duplex (2 mismatches versus consensus sequence) was used as a negative control. SiRNA transfections were performed at 4 and 40nM. Co-transfection of HIPK3-targeting siRNA duplex was tested in cells treated or not with 50ng/ml recombinant human BMP6 (A) and 5ng/ml recombinant human TGFβ (B) for 18 hours in cells pre-starved for 2 hours in serum-free culture medium. Cells were harvested 48 hours after transfection and 10µl of lysates were used for the *Dual Luciferase Assay*. Data are representative of two or three independant duplicated experiments and are presented as a ratio between firefly and renilla luciferases.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As used herein the terms "polynucleotides", "nucleic acids" and "oligonucleotides" are used interchangeably and include, but are not limited to RNA, DNA, RNA/DNA sequences of more than one nucleotide in either single chain or duplex form. The polynucleotide sequences of the present invention may be prepared from any known method including, but not limited to, any synthetic method, any recombinant method, any *ex vivo* generation method and the like, as well as combinations thereof.

The term "polypeptide" means herein a polymer of amino acids having no specific length. Thus, peptides, oligopeptides and proteins are included in the definition of "polypeptide" and these terms are used interchangeably throughout the specification, as well as in the claims. The term "polypeptide" does not exclude post-translational modifications such as polypeptides having covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like. Also encompassed by this definition of "polypeptide" are homologs thereof.

By the term "homologs" is meant structurally similar genes contained within a given species, orthologs are functionally equivalent genes from a given species or strain, as determined for example, in a standard complementation assay. Thus, a polypeptide of interest can be used not only as a model for identifying similiar genes in given strains, but also to identify homologs and orthologs of the polypeptide of interest in other species. The orthologs, for example, can also be identified in a conventional complementation assay. In addition or alternatively, such orthologs can be expected to exist in bacteria (or other kind of

cells) in the same branch of the phylogenic tree, as set forth, for example, at ftp://ftp.cme.msu.edu/pub/rdp/SSU-rRNA/SSU/Prok.phylo.

As used herein the term "prey polynucleotide" means a chimeric polynucleotide encoding a polypeptide comprising (i) a specific domain; and (ii) a polypeptide that is to be tested for interaction with a bait polypeptide. The specific domain is preferably a transcriptional activating domain.

As used herein, a "bait polynucleotide" is a chimeric polynucleotide encoding a chimeric polypeptide comprising (i) a complementary domain; and (ii) a polypeptide that is to be tested for interaction with at least one prey polypeptide. The complementary domain is preferably a DNA-binding domain that recognizes a binding site that is further detected and is contained in the host organism.

As used herein "complementary domain" is meant a functional constitution of the activity when bait and prey are interacting; for example, enzymatic activity.

As used herein "specific domain" is meant a functional interacting activation domain that may work through different mechanisms by interacting directly or indirectly through intermediary proteins with RNA polymerase II or III-associated proteins in the vicinity of the transcription start site.

As used herein the term "complementary" means that, for example, each base of a first polynucleotide is paired with the complementary base of a second polynucleotide whose orientation is reversed. The complementary bases are A and T (or A and U) or C and G.

The term "sequence identity" refers to the identity between two peptides or between two nucleic acids. Identity between sequences can be determined by comparing a position in each of the sequences which may be aligned for the purposes of comparison. When a position in the compared sequences is occupied by the same base or amino acid, then the sequences are identical at that position. A degree of sequence identity between nucleic acid sequences is a function of the number of identical nucleotides at positions shared by these sequences. A degree of identity between amino acid sequences is a function of the number of identical amino acid sequences that are shared between these sequences. Since two polypeptides may each (i) comprise a sequence (i.e., a portion of a complete polynucleotide sequence) that is similar between two polynucleotides, and (ii) may further comprise a sequence that is divergent between two polynucleotides, sequence identity comparisons between two or more polynucleotides over a "comparison window" refers to the conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference nucleotide sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less compared to the reference

sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

To determine the percent identity of two amino acids sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison. For example, gaps can be introduced in the sequence of a first amino acid sequence or a first nucleic acid sequence for optimal alignment with the second amino acid sequence or second nucleic acid sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, the molecules are identical at that position.

The percent identity between the two sequences is a function of the number of identical positions shared by the sequences. Hence % identity = number of identical positions / total number of overlapping positions X 100.

In this comparison the sequences can be the same length or may be different in length. Optimal alignment of sequences for determining a comparison window may be conducted by the local homology algorithm of Smith and Waterman (*J. Theor. Biol.,* 91 (2) pgs. 370-380 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Miol. Biol.*, 48(3) pgs. 443-453 (1972), by the search for similarity via the method of Pearson and Lipman, *PNAS, USA*, 85(5) pgs. 2444-2448 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetic Computer Group, 575, Science Drive, Madison, Wisconsin) or by inspection.

The best alignment (i.e., resulting in the highest percentage of identity over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide by nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size) and multiplying the result by 100 to yield the percentage of sequence identity. The same process can be applied to polypeptide sequences.

The percentage of sequence identity of a nucleic acid sequence or an amino acid sequence can also be calculated using BLAST software (Version 2.06 of September 1998) with the default or user defined parameter.

The term "sequence similarity" means that amino acids can be modified while retaining the same function. It is known that amino acids are classified according to the nature of their side groups and some amino acids such as the basic amino acids can be interchanged for one another while their basic function is maintained.

The term "isolated" as used herein means that a biological material such as a nucleic acid or protein has been removed from its original environment in which it is naturally present. For example, a polynucleotide present in a plant, mammal or animal is present in its natural state and is not considered to be isolated. The same polynucleotide separated from the adjacent nucleic acid sequences in which it is naturally inserted in the genome of the plant or animal is considered as being "isolated."

The term "isolated" is not meant to exclude artificial or synthetic mixtures with other compounds, or the presence of impurities which do not interfere with the biological activity and which may be present, for example, due to incomplete purification, addition of stabilizers or mixtures with pharmaceutically acceptable excipients and the like.

"Isolated polypeptide" or "isolated protein" as used herein means a polypeptide or protein which is substantially free of those compounds that are normally associated with the polypeptide or protein in a naturally state such as other proteins or polypeptides, nucleic acids, carbohydrates, lipids and the like.

The term "purified" as used herein means at least one order of magnitude of purification is achieved, preferably two or three orders of magnitude, most preferably four or five orders of magnitude of purification of the starting material or of the natural material. Thus, the term "purified" as utilized herein does not mean that the material is 100% purified and thus excludes any other material.

The term "variants" when referring to, for example, polynucleotides encoding a polypeptide variant of a given reference polypeptide are polynucleotides that differ from the reference polypeptide but generally maintain their functional characteristics of the reference polypeptide. A variant of a polynucleotide may be a naturally occurring allelic variant or it may be a variant that is known naturally not to occur. Such non-naturally occurring variants of the reference polynucleotide can be made by, for example, mutagenesis techniques, including those mutagenesis techniques that are applied to polynucleotides, cells or organisms.

Generally, differences are limited so that the nucleotide sequences of the reference and variant are closely similar overall and, in many regions identical.

Variants of polynucleotides according to the present invention include, but are not limited to, nucleotide sequences which are at least 95% identical after alignment to the reference polynucleotide encoding the reference polypeptide. These variants can also have 96%, 97%, 98% and 99.999% sequence identity to the reference polynucleotide.

Nucleotide changes present in a variant polynucleotide may be silent, which means that these changes do not alter the amino acid sequences encoded by the reference polynucleotide.

Substitutions, additions and/or deletions can involve one or more nucleic acids. Alterations can produce conservative or non-conservative amino acid substitutions, deletions and/or additions.

Variants of a prey or a SID® polypeptide encoded by a variant polynucleotide can possess a higher affinity of binding and/or a higher specificity of binding to its protein or polypeptide counterpart, against which it has been initially selected. In another context, variants can also loose their ability to bind to their protein or polypeptide counterpart.

By "fragment of a polynucleotide" or "fragment of a SID® polynucleotide" is meant that fragments of these sequences have at least 12 consecutive nucleotides, or between 12 and 5,000 consecutive nucleotides, or between 12 and 10,000 consecutive nucleotides, or between 12 and 20,000 consecutive nucleotides.

By "fragment of a polypeptide" or "fragment of a SID® polypeptide" is meant that fragments of these sequences have at least 4 consecutive amino acids, or between 4 and 1,700 consecutive amino acids, or between 4 and 3,300 consecutive amino acids, or between 4 and 6,600 consecutive amino acids.

By "anabolic pathway" is meant a reaction or series of reactions in a metabolic pathway that synthesize complex molecules from simpler ones, usually requiring the input of energy. An anabolic pathway is the opposite of a catabolic pathway.

As used herein, a "catabolic pathway" is a series of reactions in a metabolic pathway that break down complex compounds into simpler ones, usually releasing energy in the process. A catabolic pathway is the opposite of an anabolic pathway.

As used herein, "drug metabolism" is meant the study of how drugs are processed and broken down by the body. Drug metabolism can involve the study of enzymes that break down drugs, the study of how different drugs interact within the body and how diet and other ingested compounds affect the way the body processes drugs.

As used herein, "metabolism" means the sum of all of the enzyme-catalyzed reactions in living cells that transform organic molecules.

By "secondary metabolism" is meant pathways producing specialized metabolic products that are not found in every cell.

As used herein, "SID®" means a Selected Interacting Domain and is identified as follows: for each bait polypeptide screened, selected prey polypeptides are compared. Overlapping fragments in the same ORF or CDS define the selected interacting domain.

As used herein the term "PIM®" means a protein-protein interaction map. This map is obtained from data acquired from a number of separate screens using different bait polypeptides and is designed to map out all of the interactions between the polypeptides.

The term "affinity of binding", as used herein, can be defined as the affinity constant Ka when a given SID® polypeptide of the present invention which binds to a polypeptide and is the following mathematical relationship:

$$Ka = \frac{[\text{SID®/polypeptide complex}]}{[\text{free SID®}]\,[\text{free polypeptide}]}$$

wherein [free SID®], [free polypeptide] and [SID®/polypeptide complex] consist of the concentrations at equilibrium respectively of the free SID® polypeptide, of the free polypeptide onto which the SID® polypeptide binds and of the complex formed between SID® polypeptide and the polypeptide onto which said SID® polypeptide specifically binds.

The affinity of a SID® polypeptide of the present invention or a variant thereof for its polypeptide counterpart can be assessed, for example, on a Biacore™ apparatus marketed by Amersham Pharmacia Biotech Company such as described by Szabo *et al.* (*Curr Opin Struct Biol* **5** pgs. 699-705 (1995)) and by Edwards and Leartherbarrow (*Anal. Biochem* 246 pgs. 1-6 (1997)).

As used herein the phrase "at least the same affinity" with respect to the binding affinity between a SID® polypeptide of the present invention to another polypeptide means that the Ka is identical or can be at least two-fold, at least three-fold or at least five fold greater than the Ka value of reference.

As used herein, the term "modulating compound" means a compound that inhibits or stimulates or can act on another protein which can inhibit or stimulate the protein-protein interaction of a complex of two polypeptides or the protein-protein interaction of two polypeptides.

More specifically, the present invention comprises complexes of polypeptides or polynucleotides encoding the polypeptides composed of a bait polypeptide, or a bait polynucleotide encoding a bait polypeptide and a prey polypeptide or a prey polynucleotide encoding a prey polypeptide. The prey polypeptide or prey polynucleotide encoding the prey polypeptide is capable of interacting with a bait polypeptide of interest in various hybrid systems.

As described in the background of the present invention, there are various methods known in the art to identify prey polypeptides that interact with bait polypeptides of interest. These methods include, but are not limited to, generic two-hybrid systems as described by Fields *et al.* (*Nature*, 340:245-246 (1989)) and more specifically in U.S. Patent Nos. 5,283,173, 5,468,614 and 5,667,973, which are hereby incorporated by reference; the

reverse two-hybrid system described by Vidal *et al.* (*supra*); the two plus one hybrid method described, for example, in Tirode *et al.* (*supra*); the yeast forward and reverse 'n'-hybrid systems as described in Vidal and Legrain (*supra*); the method described in WO 99/42612; those methods described in Legrain *et al.* (*FEBS Letters* 480 pgs. 32-36 (2000)) and the like.

The present invention is not limited to the type of method utilized to detect protein-protein interactions and therefore any method known in the art and variants thereof can be used. It is however better to use the method described in WO99/42612 or WO00/66722, both references incorporated herein by reference due to the methods' sensitivity, reproducibility and reliability.

Protein-protein interactions can also be detected using complementation assays such as those described by Pelletier *et al.* at http://www.abrf.org/JBT/Articles/JBT0012/jbt0012.html, WO 00/07038 and WO98/34120.

Although the above methods are described for applications in the yeast system, the present invention is not limited to detecting protein-protein interactions using yeast, but also includes similar methods that can be used in detecting protein-protein interactions in, for example, mammalian systems as described, for example in Takacs *et al.* (*Proc. Natl. Acad. Sci., USA*, **90** (21):10375-79 (1993)) and Vasavada *et al.* (*Proc. Natl. Acad. Sci., USA*, 88 (23):10686-90 (1991)), as well as a bacterial two-hybrid system as described in Karimova *et al.* (1998), WO99/28746, WO00/66722 and Legrain *et al.* (*FEBS Letters*, **480** pgs. 32-36 (2000)).

The above-described methods are limited to the use of yeast, mammalian cells and *Escherichia coli* cells, the present invention is not limited in this manner. Consequently, mammalian and typically human cells, as well as bacterial, yeast, fungus, insect, nematode and plant cells are encompassed by the present invention and may be transfected by the nucleic acid or recombinant vector as defined herein.

Examples of suitable cells include, but are not limited to, VERO cells, HELA cells such as ATCC No. CCL2, CHO cell lines such as ATCC No. CCL61, COS cells such as COS-7 cells and ATCC No. CRL 1650 cells, W138, BHK, HepG2, 3T3 such as ATCC No. CRL6361, A549, PC12, K562 cells, 293 cells, Sf9 cells such as ATCC No. CRL1711 and Cv1 cells such as ATCC No. CCL70.

Other suitable cells that can be used in the present invention include, but are not limited to, prokaryotic host cells strains such as *Escherichia coli*, (e.g., strain DH5-α), *Bacillus subtilis*, *Salmonella typhimurium*, or strains of the genera of *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

Further suitable cells that can be used in the present invention include yeast cells such as those of *Saccharomyces* such as *Saccharomyces cerevisiae*.

The bait polynucleotide, as well as the prey polynucleotide can be prepared according to the methods known in the art such as those described above in the publications and patents reciting the known method *per se*.

The bait and the prey polynucleotide of the present invention is obtained from transforming growth factor β cDNA, or variants of cDNA fragment from a library of transforming growth factor β, and fragments from the genome or transcriptome of transforming growth factor β cDNA ranging from about 12 to about 5,000, or about 12 to about 10,000 or from about 12 to about 20,000. The prey polynucleotide is then selected, sequenced and identified.

A transforming growth factor β super-family of cytokines prey library is prepared from the transforming growth factor β cDNA and constructed in the specially designed prey vector pP6 as shown in Figure 3 after ligation of suitable linkers such that every cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcription activation domain of a reporter gene. Any transcription activation domain can be used in the present invention. Examples include, but are not limited to, Gal4,YP16, B42, His and the like. Toxic reporter genes, such as CAT^R, CYH2, CYH1, URA3, bacterial and fungi toxins and the like can be used in reverse two-hybrid systems.

The polypeptides encoded by the nucleotide inserts of the transforming growth factor β prey library thus prepared are termed "prey polypeptides" in the context of the presently described selection method of the prey polynucleotides.

The bait polynucleotides can be inserted in bait plasmid pB27 or pB28 as illustrated in Figure 8 and Figure 9. The bait polynucleotide insert is fused to a polynucleotide encoding the binding domain of, for example, the Gal4 DNA binding domain and the shuttle expression vector is used to transform cells.

The bait polynucleotides used in the present invention are described in Table 1.

As stated above, any cells can be utilized in transforming the bait and prey polynucleotides of the present invention including mammalian cells, bacterial cells, yeast cells, insect cells and the like.

In an embodiment, the present invention identifies protein-protein interactions in yeast. In using known methods a prey positive clone is identified containing a vector which comprises a nucleic acid insert encoding a prey polypeptide which binds to a bait polypeptide of interest. The method in which protein-protein interactions are identified comprises the following steps:

i) mating at least one first haploid recombinant yeast cell clone from a recombinant yeast cell clone library that has been transformed with a plasmid containing the prey polynucleotide to be assayed with a second haploid recombinant yeast cell clone transformed with a plasmid containing a bait polynucleotide encoding for the bait polypeptide;

ii) cultivating diploid cell clones obtained in step i) on a selective medium; and

iii) selecting recombinant cell clones which grow on the selective medium.

This method may further comprise the step of:

iv) characterizing the prey polynucleotide contained in each recombinant cell clone which is selected in step iii).

In yet another embodiment of the present invention, *in lieu* of yeast, *Escherichia coli* is used in a bacterial two-hybrid system, which encompasses a similar principle to that described above for yeast, but does not involve mating for characterizing the prey polynucleotide.

In yet another embodiment of the present invention, mammalian cells and a method similar to that described above for yeast for characterizing the prey polynucleotide are used.

By performing the yeast, bacterial or mammalian two-hybrid system, it is possible to identify for one particular bait an interacting prey polypeptide. The prey polynucleotide that has been selected by testing the library of preys in a screen using the two-hybrid, two plus one hybrid methods and the like, encodes the polypeptide interacting with the protein of interest.

The present invention is also directed, in a general aspect, to a complex of polypeptides, polynucleotides encoding the polypeptides composed of a bait polypeptide or bait polynucleotide encoding the bait polypeptide and a prey polypeptide or prey polynucleotide encoding the prey polypeptide capable of interacting with the bait polypeptide of interest. These complexes are identified in Table 2.

In another aspect, the present invention relates to a complex of polynucleotides consisting of a first polynucleotide, or a fragment thereof, encoding a prey polypeptide that interacts with a bait polypeptide and a second polynucleotide or a fragment thereof. This fragment has at least 12 consecutive nucleotides, but can have between 12 and 5,000 consecutive nucleotides, or between 12 and 10,000 consecutive nucleotides or between 12 and 20,000 consecutive nucleotides.

The complexes of the two interacting polypeptides listed in Table 2 and the sets of two polynucleotides encoding these polypeptides also form part of the present invention.

In yet another embodiment, the present invention relates to an isolated complex of at least two polypeptides encoded by two polynucleotides wherein said two polypeptides are associated in the complex by affinity binding and are depicted in columns 1 and 4 of Table 2.

In yet another embodiment, the present invention relates to an isolated complex comprising at least a polypeptide as described in column 1 of Table 2 and a polypeptide as described in column 4 of Table 2. The present invention is not limited to these polypeptide complexes alone but also includes the isolated complex of the two polypeptides in which

fragments and/or homologous polypeptides exhibit at least 95% sequence identity, as well as from 96% sequence identity to 99.999% sequence identity.

Also encompassed in another embodiment of the present invention is an isolated complex in which the SID® of the prey polypeptides encoded by SEQ ID N°27 to 64 in Table 3 form the isolated complex.

Besides the isolated complexes described above, nucleic acids coding for a Selected Interacting Domain (SID®) polypeptide or a variant thereof or any of the nucleic acids set forth in Table 3 can be inserted into an expression vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. Such transcription elements include a regulatory region and a promoter. Thus, the nucleic acid which may encode a marker compound of the present invention is operably linked to a promoter in the expression vector. The expression vector may also include a replication origin.

A wide variety of host/expression vector combinations are employed in expressing the nucleic acids of the present invention. Useful expression vectors that can be used include, for example, segments of chromosomal, non-chromosomal and synthetic DNA sequences. Suitable vectors include, but are not limited to, derivatives of SV40 and pcDNA and known bacterial plasmids such as col EI, pCR1, pBR322, pMal-C2, pET, pGEX as described by Smith et al (1988), pMB9 and derivatives thereof, plasmids such as RP4, phage DNAs such as the numerous derivatives of phage I such as NM989, as well as other phage DNA such as M13 and filamentous single stranded phage DNA; yeast plasmids such as the 2 micron plasmid or derivatives of the 2m plasmid, as well as centomeric and integrative yeast shuttle vectors; vectors useful in eukaryotic cells such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or the expression control sequences; and the like.

For example in a baculovirus expression system, both non-fusion transfer vectors, such as, but not limited to pVL941 (*Bam*HI cloning site Summers), pVL1393 (*Bam*HI, *Sma*I, *Xba*I, *Eco*RI, *Not*I, *Xma*III, *Bgl*II and *Pst*I cloning sites; Invitrogen), pVL1392 (*Bgl*II, *Pst*I, *Not*I, *Xma*III, *Eco*RI, *Xba*I, *Sma*I and *Bam*HI cloning site; Summers and Invitrogen) and pBlue*Bac*III (*Bam*HI, *Bgl*II, *Pst*I, *Nco*I and *Hind*III cloning site, with blue/white recombinant screening, Invitrogen), and fusion transfer vectors such as, but not limited to, pAc700 (*Bam*HI and *Kpn*I cloning sites, in which the *Bam*HI recognition site begins with the initiation codon; Summers), pAc701 and pAc70-2 (same as pAc700, with different reading frames), pAc360 (*Bam*HI cloning site 36 base pairs downstream of a polyhedrin initiation codon; Invitrogen (1995)) and pBlueBacHisA, B, C (three different reading frames with *Bam*HI, *Bgl*II, *Pst*I, *Nco*I and *Hind*III cloning site, an N-terminal peptide for ProBond purification and blue/white recombinant screening of plaques; Invitrogen (220) can be used.

Mammalian expression vectors contemplated for use in the invention include vectors with inducible promoters, such as the dihydrofolate reductase promoters, any expression vector with a DHFR expression cassette or a DHFR/methotrexate co-amplification vector such as pED (*Pst*I, *Sal*I, Sba I, Sma I and *Eco*RI cloning sites, with the vector expressing both the cloned gene and DHFR; Kaufman, 1991). Alternatively a glutamine synthetase/methionine sulfoximine co-amplification vector, such as pEE14 (*Hind*III, *Xba*I, *Sma*I, *Sba*I, *Eco*RI and *Bcl*I cloning sites in which the vector expresses glutamine synthetase and the cloned gene; Celltech). A vector that directs episomal expression under the control of the Epstein Barr Virus (EBV) or nuclear antigen (EBNA) can be used such as pREP4 (*Bam*HI, *Sfi*I, *Xho*I, *Not*I, *Nhe*I, *Hind*III, *Nhe*I, *Pvu*II and *Kpn*I cloning sites, constitutive RSV-LTR promoter, hygromycin selectable marker; Invitrogen), pCEP4 (*Bam*HI, *Sfi*I, *Xho*I, *Not*I, *Nhe*I, *Hind*III, *Nhe*I, *Pvu*II and *Kpn*I cloning sites, constitutive hCMV immediate early gene promoter, hygromycin selectable marker; Invitrogen), pMEP4 (*Kpn*I, *Pvu*I, *Nhe*I, *Hind*III, *Not*I, *Xho*I, *Sfi*I, *Bam*HI cloning sites, inducible methallothionein IIa gene promoter, hygromycin selectable marker, Invitrogen), pREP8 (*Bam*HI, *Xho*I, *Not*I, *Hind*III, *Nhe*I and *Kpn*I cloning sites, RSV-LTR promoter, histidinol selectable marker; Invitrogen), pREP9 (*Kpn*I, *Nhe*I, *Hind*III, *Not*I, *Xho*I, *Sfi*I, *Bam*HI cloning sites, RSV-LTR promoter, G418 selectable marker; Invitrogen), and pEBVHis (RSV-LTR promoter, hygromycin selectable marker, N-terminal peptide purifiable via ProBond resin and cleaved by enterokinase; Invitrogen).

Selectable mammalian expression vectors for use in the invention include, but are not limited to, pRc/CMV (*Hind*III, *Bst*XI, *Not*I, *Sba*I and *Apa*I cloning sites, G418 selection, Invitrogen), pRc/RSV (*Hind*II, *Spe*I, *Bst*XI, *Not*I, *Xba*I cloning sites, G418 selection, Invitrogen) and the like. Vaccinia virus mammalian expression vectors (see, for example Kaufman 1991 that can be used in the present invention include, but are not limited to, pSC11 (*Sma*I cloning site, TK- and β-gal selection), pMJ601 (*Sal*I, *Sma*I, *Afl*I, *Nar*I, *Bsp*MII, *Bam*HI, *Apa*I, *Nhe*I, *Sac*II, *Kpn*I and *Hind*III cloning sites; TK- and β-gal selection), pTKgptF1S (*Eco*RI, *Pst*I, *Sal*II, *Acc*I, *Hind*II, *Sba*I, *Bam*HI and *Hpa* cloning sites, TK or XPRT selection) and the like.

Yeast expression systems that can also be used in the present include, but are not limited to, the non-fusion pYES2 vector (*Xba*I, *Sph*I, *Sho*I, *Not*I, *Gst*XI, *Eco*RI, *Bst*XI, *Bam*HI, *Sac*I, *Kpn*I and *Hind*III cloning sites, Invitrogen), the fusion pYESHisA, B, C (*Xba*I, *Sph*I, *Sho*I, *Not*I, *Bst*XI, *Eco*RI, *Bam*HI, *Sac*I, *Kpn*I and *Hind*III cloning sites, N-terminal peptide purified with ProBond resin and cleaved with enterokinase; Invitrogen), pRS vectors and the like.

CLAIMS

What is claimed is:

1. A complex between two interacting proteins as defined in columns 1 and 4 in Table 2.
2. A complex between two polynucleotides encoding for the polypeptides of claim 1.
3. A recombinant host cell expressing the interacting polypeptides of said complex of protein-protein interaction of claim 1.
4. Use of a SID®, an interaction or a prey to screen molecules that inhibit TGFβ or inhibit a TGFβ super-family of cytokines pathway.
5. A molecule that inhibits inhibit TGFβ or inhibits a TGFβ super-family of cytokines pathway.
6. Use according to Claim 4, wherein said screening occurs in mammalian cells or yeast cells.
7. A SID® polypeptide comprising the SEQ ID No 63 to 98.
8. A SID® polynucleotide comprising the SEQ ID No 27 to 62.
9. A vector comprising the SID® polynucleotide comprising the SEQ ID No 27 to 62.
10. A fragment of said SID® polypeptide according to Claim 7.
11. A variant of said SID® polypeptide according to Claim 7.
12. A fragment of said SID® polynucleotide according to Claim 8.
13. A variant of said SID® polynucleotide according to Claim 8.
14. A vector comprising the SID® polynucleotide according to any one of Claims 8, 12 or 13.
15. A recombinant host cell containing the vectors according to Claim 14.
16. A pharmaceutical composition comprising the molecule of claim 5 and a pharmaceutically acceptable carrier.
17. A pharmaceutical composition comprising a SID® polypeptide SEQ ID No 63 to 98 and a pharmaceutically acceptable carrier.
18. A pharmaceutical composition comprising the recombinant host cells of Claim 15 and a pharmaceutically acceptable carrier.
19. A protein chip comprising the polypeptides of Table 2.
20. Use of a ZNF8 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

21. Use of a LAPTm5 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

22. Use of a RNF11 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

23. Use of a LMO4 protein for the preparation of a medicament for treating prostate cancer.

24. Use of a PPC1 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

25. Use of an HYPA protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

26. Use of a PTP protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

27. Use of an HYPK3 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

28. Use of a KIAA1196 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

29. Use of a FL20037 protein for the preparation of a medicament for treating diseases and /or disorders linked or involving a TGFβ super-family of cytokines.

30. Use of a complex between two interacting proteins as defined in columns 1 and 4 in table 2 to screen I molecules for diagnosis or treating transforming growth factor β disorders and/or diseases.